

Association for Information Systems AIS Electronic Library (AISeL)

ECIS 2013 Completed Research

ECIS 2013 Proceedings

7-1-2013

Predicting And Economically Exploiting Utility Thresholds With Utility-Based Recommendation Systems

Verena Dorner

University of Passau, Passau, Germany, verena.dorner@uni-passau.de

Michael Scholz

University of Passau, Passau, Germany, michael.scholz@uni-passau.de

Follow this and additional works at: http://aisel.aisnet.org/ecis2013_cr

Recommended Citation

Dorner, Verena and Scholz, Michael, "Predicting And Economically Exploiting Utility Thresholds With Utility-Based Recommendation Systems" (2013). *ECIS 2013 Completed Research*. 30.

http://aisel.aisnet.org/ecis2013_cr/30

This material is brought to you by the ECIS 2013 Proceedings at AIS Electronic Library (AISeL). It has been accepted for inclusion in ECIS 2013 Completed Research by an authorized administrator of AIS Electronic Library (AISeL). For more information, please contact elibrary@aisnet.org.

PREDICTING AND ECONOMICALLY EXPLOITING UTILITY THRESHOLDS WITH UTILITY-BASED RECOMMENDATION SYSTEMS

Dorner, Verena, University of Passau, Innstr. 43, 94032 Passau, Germany,
verena.dorner@uni-passau.de

Scholz, Michael, University of Passau, Innstr. 43, 94032 Passau, Germany,
michael.scholz@uni-passau.de

Abstract

Consumers spend a lot of time and effort on online product search in order to find the best match to their preferences. Recommendation systems promise to decrease these search costs. Much recent work has focused on refining methods for finding the best alternatives for a consumer. The question of how many of these alternatives the consumer actually wants to see has remained largely unanswered. This paper proposes utility thresholds as a novel approach to identifying the optimal recommendation set size. Beyond improving recommendation systems, utility thresholds improve business decision support by enabling willingness-to-pay estimation and individually optimal product configurations. Our empirical evaluation shows that utility threshold prediction is better for factors related to the recommendation process than personal factors. Search costs are reduced and willingness-to-pay estimates improved.

Keywords: Recommender Systems, E-Commerce, Economic Implications.

1 Introduction

Recommendation systems (RS) became popular after the number of products available online had begun to exceed the information processing capabilities of consumers (Resnick and Varian 1997, Schafer et al. 2001). A number of different recommendation approaches have since been explored and successfully applied in practice, e.g. on Amazon.com (Burke 2002, Herlocker et al. 2004, De Bruyn 2008). But despite their popularity, little research has been done to investigate how RSs can be leveraged to improve estimates of economically and managerially relevant data. This paper examines utility-based RSs, which generate recommendations by measuring consumers' utility functions and predicting individual product utilities. They elicit preferences more accurately than other approaches – an advantage which has made them popular with researchers (Ansari et al. 2000, Huang 2011, Scholz and Dorner 2012) and has aroused practitioners' interest (e.g. Dell's Computer Advisor). More accurate preference estimates – aside from benefitting consumers by generating more fitting recommendations – open up new opportunities for improving management decision support. Specifically, we show how utility-based RSs can be used for estimating consumers' willingness to pay and for identifying profit-maximizing product configurations. Our study is one of the first to examine utility-based RSs from this angle.

The major challenge for any RS is separating attractive products – for which the willingness to pay is greater than zero – from unattractive products. The solution we suggest is identifying each consumer's utility threshold. It indicates the product with the lowest utility that she still finds attractive. Combined with her utility function for the product, the utility threshold enables us to estimate her willingness to pay and to find out which attribute combinations she would purchase. In this paper, we examine two distinct sets of threshold predictors, personal and recommendation process related factors. Our research helps design RSs that “understand” consumer preferences better and require much less effort on part of the consumer than traditional methods like choice-based conjoint analysis; consumer search and evaluation costs decrease.

Section 2 describes the theoretical foundations of our study. Section 3 presents our research model and Section 4 the empirical evaluation and results. Section 5 discusses the applications and economic implications of our findings in detail. Section 6 summarizes our study and its limitations.

2 Theoretical Foundations

RSs are “software systems that elicit the interests or preferences of individual users for products, either explicitly or implicitly, and make recommendations accordingly” (Xiao and Benbasat 2007). Common approaches to finding recommendable products for consumers are collaborative and content-based filtering (Burke 2002, Xiao and Benbasat 2007). Content-based RSs use customers' past purchasing histories to generate recommendations. Collaborative filtering systems recommend products to customers with similar purchasing histories. Unfortunately, both approaches only work well if a lot of (historical) data on customers is already available (*cold start problem*): new or anonymous customers cannot be provided with useful recommendations (Adomavicius and Tuzhilin 2005).

Utility-based RSs do not suffer from the cold start problem because they do not generate recommendations based on historical data, but based on revealed attribute preferences (Huang 2011). Asking consumers to state their preferences prior to purchase protracts the recommendation process but leads to more accurate preference and utility estimates (Ansari et al. 2000). Consumers benefit from improved recommendation quality; companies benefit from applying this information to improve managerial decision making, e.g. deriving consumers' willingness to pay to improve pricing decisions (Jedidi et al. 1996, Miller et al. 2011).

Utility-based RSs support consumers' "natural" decision processes (Hauser and Wernerfelt 1990). When faced with a preferential choice task with many options, like searching for "the best" product, consumers usually try to reduce their cognitive processing costs by following a two-stage decision process (Häubl and Trifts 2000). In the screening stage, all products outside the consumer's attribute aspiration intervals (e.g. zero to maximal price) are discarded. In the evaluation stage, consumers evaluate the remaining products – their consideration set – in detail (Hauser and Wernerfelt 1990), and finally they choose the product which fits their preferences best.

For generating recommendations, utility-based RSs require one set of inputs for each stage. In the screening stage, the consumer must provide the aspiration intervals for each product attribute. In the evaluation stage, she must state her preferences for each product attribute (single-attribute utility functions). The RS then combines these single-attribute utility functions into one multi-attribute utility function to compute overall product utilities, based on multiple attribute utility theory (MAUT; Wallenius et al. 2008). All products can now be displayed in descending order of their expected utilities. (Cognitive) costs for both screening products and evaluating the consideration set are thus decreased.

But depending on how broadly the consumer defines her aspiration intervals, she may still find herself faced with a very long list of products. In such situations, the consumer is liable to grow uncertain and lose confidence in her decisions; choice deferral becomes more likely (Xiao and Benbasat 2007, White and Hoffrage 2009). We suggest that these issues can be circumvented by presenting not all products but only her consideration set to a consumer because only the products in the consideration set are sufficiently attractive for her to really consider purchasing (Andrews and Srinivasan 1995, Jedidi et al. 1996). Among the products which pass the screening stage, some products may not really be attractive. Although their attribute levels lie within the aspiration intervals, expected evaluation costs outweigh expected utilities: they fall below the utility threshold. Estimating consideration set size is thus equivalent to estimating the utility of the last product that is just attractive enough for the consumer to evaluate in detail (Jedidi et al. 1996). Once we know a consumer's utility threshold, we can also compute her willingness to pay. In the following, we outline two approaches to estimating the utility threshold.

The first approach uses certain personal and demographic consumer characteristics which are routinely available to online shops and retailers. Hence, it is particularly interesting to find out whether this information can be leveraged for predicting utility thresholds. One demographic factor which may systematically influence product utility assessments is age, which has been shown to have a positive effect both on consideration set size and purchasing probability (Frag et al. 2003). Another such factor is gender (e.g. Darley and Smith 1995, Mitchell and Walsh 2004, Karjaluoto et al. 2005). Male consumers are more likely to engage in variety-seeking purchasing (Helmig 1997), display lower levels of brand involvement (Guest 1964), and are less likely to perceive product risk (Darley and Smith 1995) than female consumers. Female consumers are more likely to be *recreational* shoppers; male consumers are more time-conscious and more likely to be *satisficers*: they are content to find a satisfactory but not necessarily the perfect product (Mitchell and Walsh 2004). Product (category) expertise has been shown repeatedly to have a significant effect on attribute perceptions and thus product utility assessments, i.e. on consideration set content. Sambandam and Lord (1995) argue that expertise positively affects product knowledge and choice satisfaction, which in turn influence consideration set size, but evidence on the direction of their influence is contradictory (e.g. Chernev et al. 2003, Frag et al. 2003, Gilbride and Allenby 2004, Karjaluoto et al. 2005).

The second approach is based on research in decision analysis (Butler et al. 2001). Although decision analytic procedures are usually aimed at finding the one best product, they can be adapted to approximate the consideration set (Scholz and Dorner 2012). In the utility exchange approach, consumers are asked to specify the utility difference between two products in terms of attribute units, e.g. costs. The revealed unit-based difference is then converted into a utility difference and used for computing threshold and consideration set (Butler et al. 2001). Another way of identifying attractive products is to interpret them as *signals* and unattractive products as *noise*, where the quality of the

outcome (i.e. correct consideration set identification) depends both on task difficulty (i.e. product similarity) and on the consumer's ability to discern signal and noise (Tanner and Swets 1954). It has yielded very good empirical results (Scholz and Dorner 2012) and appears to be more comprehensible to consumers than the abovementioned utility exchange approach. Both decision analytic approaches require additional user input, hence greater cognitive effort, for estimating the utility threshold and are very complex to implement. We suggest using more easily measured process-related factors as predictors (Section 3).

Decision analytic approaches do not require that consumers divulge personal information like age or gender. Instead, they must invest time and effort in specifying their preferences. Consumers may, however, not always be prepared to do so, and online retailers are likely to wish to be able to provide recommendations to all their customers, including "impatient" ones. We therefore decided to investigate whether it is possible to predict utility thresholds using "simple" personal characteristics as well as when using process-related characteristics.

3 Research Model

Building the RS, the first assumption is that consumer decision processes follow the two-stage *screen and evaluate* model (Hauser and Wernerfelt 1994). We also assume that 1) there exists an individual utility threshold τ , 2) products in the consideration set are sorted in descending order of their expected utilities $E[u(X_1)] \geq E[u(X_2)] \geq \dots \geq E[u(X_k)]$, and 3) the consumer's product rating permits us to infer the level of the utility threshold.

Consumers carry out detailed evaluations of the products with the highest expected utilities $E[u(X_k)]$. Products with utilities below the individual threshold τ are not attractive enough to warrant further detailed evaluation. As long as $E[u(X_k)] \geq \tau$ holds, consumers inspect products in descending order of their utilities.

Decision quality improves when RSs present their results in this order (Diehl et al. 2003, Aksoy et al. 2006). The consumer's product ratings $\phi(X_k)$ permit us to infer the position of her utility threshold and her consideration set size. Products with positive ratings are likely to be part of the consideration set; products with negative ratings are probably unattractive. The greater the extent to which the best product – as recommended by the system – corresponds to the consumer's preferences, the more likely it is that she will want to inspect the next best product(s). We propose that the consumer's rating of the best product $\phi(X_1)$ and the difference between the expected utility of the best product $E[u(X_1)]$ and the utility threshold τ are positively correlated.

Hypothesis 1: Consideration set size increases with the recommended best product's rating.

Chernev (2003) argues that decision processes are contingent on the "degree of articulation of the readily available preferences". Consumers without an available "ideal point" need to start from scratch, i.e. they must first construct their ideal attribute combination (Bettman et al. 1998) before they can begin searching for the best match. Consumers with an available ideal point presumably face a less demanding choice task. Our RS first asks consumers to perform attribute-based screening, i.e. to specify upper and lower bounds for each attribute to indicate the aspiration intervals. Consumers with an available ideal point can articulate their preferences clearly and are thus able to restrict attribute levels "around" their respective ideal levels. All products recommended to them will be "distributed" around their ideal attribute combination. These consumers will wish to see many of the remaining alternatives: they expect the displayed alternatives to have high utilities. Arguing from the cost-utility point of view, the trade-off between the cost and the potential benefit of evaluating an additional product (i.e. including it in the consideration set) is favourable towards adding more products. The expected gain in utility outweighs the expected increase in cost. We therefore expect that consumers who restrict attribute levels strongly will have a larger consideration set.

Hypothesis 2a: Consideration set size increases with attribute interval restriction intensity.

However, we only expect this to hold true if the best recommended product is judged to be a good option by the consumer. Consumers who “know what they want”, and consequently restrict attribute intervals strongly, will only be prepared to evaluate many products if they expect the recommendations to be good. *Good* in this context means that they are close to the best recommended product, assuming that this product really is a good match to consumer preferences. If the first product is not a good match, consumers will be very unwilling to evaluate many more products.

Hypothesis 2b: Attribute interval restriction intensity moderates the effect of “best product rating”.

When consumers cannot draw from expertise or prior experience in a product category, they must first construct their preferences and are usually less able to specify aspiration intervals (Chernev 2003). They incur greater costs during the search process because they need to solve a more complex task (preference building plus product search and evaluation). The costs of evaluating an additional product are likely to be quite high because the consumer is incurring preference-building costs at the same time. We therefore expect less experienced consumers to prefer smaller consideration sets.

Hypothesis 3: Consumers with a lot of product expertise have larger consideration sets than consumers with little expertise.

Since male consumers are more time-conscious and more likely to be satisficers than female consumers, who tend to be recreational shoppers (Mitchell and Walsh 2004), we expect that female consumers are willing to spend time on evaluating more products in detail than their male counterparts.

Hypothesis 4: Male consumers have smaller consideration sets than female consumers.

Farag et al. (2003) showed that age influences consideration set size and purchase probabilities. They suggest that older consumers may have smaller consideration sets because they have fewer resources, in particular less time, at their disposal for shopping activities. By reducing consideration set size, they also reduce time and effort required for product evaluation.

Hypothesis 5: Older consumers have smaller consideration sets than younger consumers.

Figure 1 summarizes our research model.

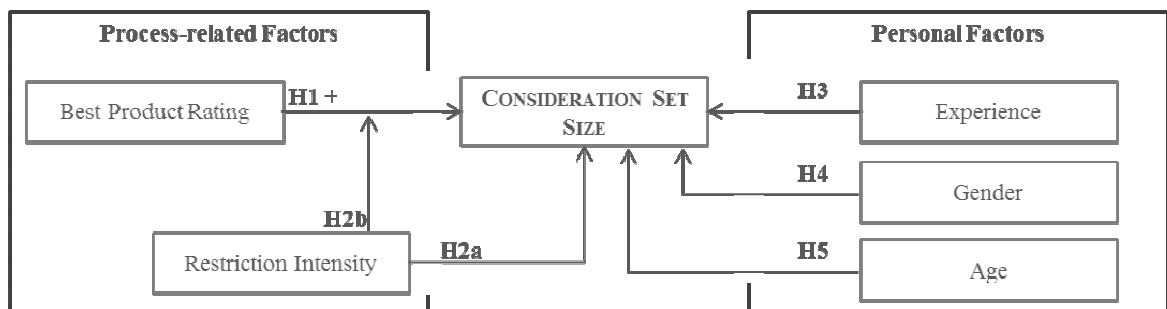


Figure 1. Research Model.

4 Empirical Investigation

4.1 Procedure

We conducted a laboratory experiment to investigate how well process-related and personal factors are suited for predicting utility thresholds¹. The participants were told to search for a new digital camera with the help of a RS which contained 105 different digital cameras described by six attributes: photo resolution, video resolution, zoom factor, display size, weight and price.

Participants were asked to complete five tasks in the experiment (Table 1). In the first task (A in Table 1), participants specified acceptable aspiration intervals for each of the six displayed product attributes; e.g. for zoom factor a minimal aspiration level of 8 and a maximal level of 12 megapixels. In the second task B, participants ranked the attributes in order of their importance, e.g. price over zoom factor over weight and so on.

Based on the revealed aspiration levels (task A), our RS generated a set of 18 stimuli (i.e. digital cameras), using three different attribute levels (minimum, maximum, and the average of these levels) for 6x3 D-optimal conjoint design. The RS presented the attributes for the stimuli set in order of the revealed attribute weights (task B).

In task C, participants ranked the stimuli in order of their perceived attractiveness. The RS then computed single-attribute utility functions and product utilities with an ordinary least squares estimator and displayed the product with the highest utility and 10 randomly selected products. In task D, participants rated these products on a 7-point scale, ranging from 1 (unattractive) to 7 (very attractive).

Finally, participants completed a survey (task E) on three personal characteristics: experience with digital cameras, age, and gender.

	Screening	Evaluation		Consideration Set	Personal Data
Task	A	B	C	D	E
System	Select Products Within Aspiration Intervals	Present Attributes in Random Order	Generate 6x3 D-optimal Stimuli	Generate Product List	–
User	Restrict Attributes to Aspiration Intervals	Rank Attributes	Rank Stimuli	Rate Products	Reveal Personal Data
Method	Conjunctive Model	Rank-Ordered Centroids	Ranking-based Conjoint Analysis based on OLS	Rating	Survey

Table 1. Experimental procedure.

¹ In Scholz and Dorner (2012), the experiment was used to evaluate two recommendation agents implementing an adapted utility exchange approach (Butler et al. 2001) and a signal detection theory-based approach.

4.2 Sample

We conducted one-on-one pre-tests with 8 students who did not take part in the final experiment. All suggestions made by at least 2 participants were implemented.

93 students from the University of Passau were invited to a lab and given instructions how to proceed. Each participant was paid 7 euros. 66% of the participants were female and the average age was 23 years, ranging from 19 to 36.

Predicting a consumer's utility threshold with the process-related factors *rating of best expected product* and *restriction intensity* requires reliable estimates for her utility functions. We performed F-tests to determine the internal consistency of the elicited utility functions, i.e. whether participants solved tasks A, B and C consistently. We excluded those who did not, which led to a final sample of 80 participants.

4.3 Preliminary Data Analysis

Before testing our research model (Section 4.4), we established that our utility estimates and utility threshold estimates were reliable and satisfactory from the consumers' point of view (Table 2).

Utility Elicitation	
First-choice hit rates	0.71
"benchmark value" (De Bruyn et al. 2008)	0.39 to 0.50
Rank correlations	0.57
"benchmark value" (De Bruyn et al. 2008)	0.48 to 0.59
Utility Thresholds	
Precision	0.77
Recall	0.81
F-measure	0.79

Table 2: Predictive validity of utility and threshold estimates

Average R^2 of 0.90 (adj. $R^2=0.84$) for the utility parameter estimates – price, photo resolution and zoom emerged as the most important attributes – indicates reliable utility measurements but not necessarily high levels of consumer satisfaction with the constructed utility functions. We measured predictive validity by comparing the predicted product ranks with observed product ratings. The results were very satisfactory, especially when compared to other utility-based RSs (Table 2).

Ratings of at least 4 out of 7 were interpreted as positive (0) and lower ratings as negative (1). Each participant's τ value, which separates attractive (0) from unattractive (1) products, was then estimated in a logistic regression. We examined τ by means of evaluating precision, recall, and F-measures. Precision measures the fraction of products in the recommendation set that were rated "attractive" (0.77). Recall indicates the fraction of attractive products in the consideration set (0.81). The F-measure expresses the balance between precision and recall, and reached the outstanding level of 0.79. The estimated threshold values are evidently suitable for predicting whether a product is attractive for a consumer (Scholz and Dorner 2012).

4.4 Results

We normalized the estimated utility thresholds over the interval [0; 1], which is bounded by the utilities of the worst and best products. Since the thresholds follow a truncated Gaussian distribution, we used Tobit regression to examine the effects of personal and process-related factors. We estimated three regression models (Table 3). The first model contains only the personal factors *age*, *gender* and *experience*; the second model only the process-related factors *rating of best expected product* and *restriction intensity*; and the third model all factors from both categories.

Model comparison (Table 3) shows that personal factors (model 1) are bad predictors for utility threshold and consideration set size. They reduce the estimation power (measured by AIC and BIC) of process-related factors (model 3). Process-related factors alone (model 2) perform much better.

	Model 1		Model 2		Model 3	
	Est.	Std. Err.	Est.	Std. Err.	Est.	Std. Err.
Intercept	0.290	0.241	1.904***	0.362	1.621***	0.474
Age	0.017	0.011	-	-	0.007	0.011
Gender	0.084	0.054	-	-	0.072	0.052
Experience	-0.001	0.033	-	-	-0.001	0.031
Rating Best Product (RBP)	-	-	-0.193**	0.066	-0.176**	0.067
Restriction Intensity (RI)	-	-	-2.560**	0.852	-2.349**	0.864
RBP x RI	-	-	0.402**	0.154	0.367*	0.156
Log-Likelihood	-3.437***		0.862***		2.114*	
AIC	16.875		8.276		11.772	
BIC	28.785		20.186		30.828	

Table 3. Model comparison results (* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$).

Higher ratings for the best expected product indicate lower utility thresholds and larger consideration sets (H1). Higher restriction intensity for attribute levels also indicates lower utility thresholds and larger consideration sets (H2a). In addition, restriction intensity moderates the relationship between best expected product rating and utility threshold (H2b). The effects of all process-related variables are illustrated in Figure 2. None of the personal factors contributed to model fitness; we therefore reject hypotheses 3, 4 and 5.

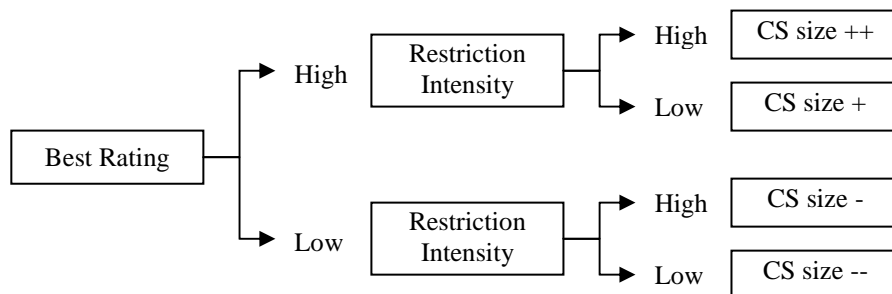


Figure 2. Interaction effect between best rating and restriction intensity.

We conducted a simulation with 25 runs to determine the predictive quality of the best model, model 2. In each run, three fourth of the data were used for regression and one fourth for prediction. The predicted thresholds were then used to measure precision, recall, and the F-measure. In relation to the values for the optimal threshold (normed threshold which discriminates best between attractive and unattractive products), we obtained remarkably high values for all indicators (Table 4). The predicted thresholds are distributed with a mean of 0.72 and a standard deviation of 0.10.

	Simulated thresholds	Optimal thresholds
Precision	0.66	0.77
Recall	0.76	0.81
F-measure	0.71	0.79

Table 4. Predictive quality of model 2.

5 Implications

Knowing a consumer's utility threshold, we can 1) reduce search and evaluation costs, 2) estimate individual willingness to pay, and 3) develop efficient product configurations. We discuss these three implications briefly in the following subsections. Both the estimation of individual willingness to pay (WTP) and the compilation of efficient product configurations are based on the well-known utility exchange approach (Butler et al. 2001).

5.1 Search and Evaluation Costs

High cognitive load can lead to a higher probability of choice deferral (White and Hoffrage 2009). Consumers' cognitive load during the online shopping process can be reduced – and sales increased – by shrinking the consideration set to the smallest possible number of products. Each product in the consideration set is evaluated, and each evaluation incurs cognitive costs, which can be measured in terms of elementary information process (EIP) steps (Johnson and Payne 1985). Two examples for EIP steps are “reading” an attribute level and “comparing” two attribute levels. The total cognitive cost of evaluation for a product set depends on the consumer's decision strategy (Pfeiffer 2012). Consumers who use a weighted additive decision strategy for computing the utility of each considered product require $3m$ EIP steps for assigning to each attribute m an importance weight. If n products are considered, $(4m+3(m-1)+1)n$ EIP steps are needed for assigning attribute values and computing product utilities. Finding the alternative with the highest overall utility requires additional $4(n-1)$ EIP steps (Pfeiffer 2012).

Consider a consumer who evaluates 6 digital cameras² described by 8 attributes. If she uses a weighted additive decision strategy to select the best product, she will incur cognitive costs of 366 EIP steps. Removing just one product (56 EIP steps) from the consideration set would correspond to an 18% decrease in cognitive costs. If we can shrink the consideration set by applying our utility threshold estimation approach, we will reduce the effort of online product search and help online retailers increase sales.

² Most consumers evaluate only up to 5 products in detail (Gu et al. 2011, Häubl and Trifts 2000).

5.2 Willingness to Pay

Both in theory and in practice, pricing strategies are defined based on assumptions about consumers' WTP and their product utilities. Researchers and practitioners alike often do not have reliable evidence for estimating either, which renders many advanced pricing models unusable. This is a problem which we propose to resolve with our RS. A consumer's WTP for a specific product is the price at which the consumer is indifferent between buying and not buying the product (Gensler et al. 2012). The utility threshold τ defines the utility level for which a consumer is indifferent between buying and not buying the product. The price of a product with a utility of τ is therefore equal to the consumer's WTP (Jedidi and Zhang 2006). Most products' utilities, however, are different from τ . The utility exchange approach (Butler et al. 2001) proposes that we can increase (decrease) the price of product X and it will remain attractive for the consumer as long as the utility of X is at least equal to τ .

Consider a consumer with a utility threshold of $\tau = 0.9$ and a utility for product X of $u(X) = 1.0$. We can increase the price of X such that the utility of X decreases to 0.9. If the price of X is 100 euros and the single-attribute utility function for price is given by $u(\text{price}) = -0.005\text{price}$, then increasing the price by 20 euros equals a decrease in utility by 0.1 units. Hence, WTP is 120 euros – in other words, the utility of product X equals τ at a price of 120 euros.

5.3 Product Configuration

When planning their product assortments, retailers usually try to balance the goals *market share growth* and *margin maximization*. They search for bundles of attribute levels which they can procure at a low price and resell at a high price. Profit-optimal product configuration requires in-depth knowledge of customers' utility functions, their utility thresholds, and the purchase prices for the possible product configurations.

Consider a consumer planning to purchase a new notebook who recently used our utility-based RS. Using her revealed preference information, we can estimate her single-attribute utility (SAU) functions for the notebook attributes processor speed, memory size, HDD size and price (Table 3).

Attribute	Minimal Level	Maximal Level	Weight (0-10)	SAU Function
Processor Speed (in GHz)	2.0	3.6	6	$-1.25+0.625x$
Memory Size (in GB)	2	16	5	$-0.143+0.071x$
HDD Size (in GB)	250	1000	7	$-0.333+0.0013x$
Price (in Euro)	400	1000	8	$1.667-0.0017x$

Table 5. Utility functions of exemplary consumer.

Let us assume that the consumer's utility threshold τ is 12 and that the retailer offers a notebook with 2.8 GHz processor speed, 4 GB memory size and 500 GB HDD for 600 euros. This notebook's utility ($U(X) = 11.1$) lies below the utility threshold. To attract this particular consumer, the retailer can choose one of four options – upgrade processor, memory or HDD, or else decrease the price to 533 euros³. At this price, the consumer would be indifferent between buying and not buying the notebook. If the purchase price were set to 450 euros, the retailer would make a profit of 83 euros (Table 4). If, however, processor speed were upgraded instead, profit would rise to 100 euros. In addition, the

³ For this particular consumer, notebook utility equals 12 (utility units) at a price of 533 Euro.

upgrade option would increase the consumer's utility to 12.6 units, exceeding her utility threshold. She would actually prefer the upgraded notebook to the cheaper notebook.

Attribute	Actual Level	Actual Purchase Price	Upgrade Option	Utility (<i>Upgraded</i>)	Purchase Price (<i>Upgraded</i>)	Profit (<i>Upgraded</i>)
Processor Speed	2.8 GHz	150 EUR	3.2 GHz	12.600	200 EUR	100 EUR
Memory Size	4 GB	150 EUR	8 GB	12.520	250 EUR	50 EUR
HDD Size	500 GB	150 EUR	750 GB	13.375	220 EUR	80 EUR
Price	600 EUR	-	533 EUR	12.000	-	83 EUR

Table 6. Upgrade opportunities.

6 Conclusion

In this paper, we propose a new approach for estimating consumers' utility thresholds with a utility-based RS. Utility thresholds determine size and content of individual recommendation sets: size increases with the number of products whose utilities surpass the threshold. Knowledge of a consumer's utility threshold has three distinct advantages. First, the RS can identify the most attractive products with greater accuracy, thus helping consumers to save cognitive costs in the search process. Second, retailers and product manufacturers can estimate consumers' individual willingness to pay for a particular product. Third, they can identify and offer profit-optimal individual product configurations.

We built one model for threshold prediction with personal variables – which influence consideration set content – and another model with process variables – which relate to preference-revealing consumer behaviour during the recommendation process. Both sets of variables do not require consumers to expend much, if any, additional effort during the recommendation process. Comparison of the two models showed that process-related variables are good predictors for utility thresholds; personal variables are not. We therefore suggest extending utility-based RSs (like Dell's computer advisor) to include measures for attribute restriction intensity and the rating for the best expected product.

Our suggested approach would benefit from, and could be extended by, further research in two areas in particular. For one, the measurement model for utility threshold estimation assumes reliable utility functions. Many of the methods recently proposed in operations research and marketing for measuring utility functions more reliably, however, require a lot of consumer effort (Huang 2011). Integrating our approach with other utility-based recommendation techniques such as radial basis function networks or SMARTER – these two approaches generate particularly accurate results (Lin et al. 2005, Huang 2011) – could reduce preference elicitation and product evaluation effort, while preserving high recommendation quality.

For another, the cognitive processes of consumers during utility and price evaluations are assumed to be similar. There is some evidence, however, that “utility units” and “price units” might be perceived differently by the consumer and cannot always be used interchangeably (Scholz and Dorner 2012). Further research into how consumers process and evaluate utility and price information could help improve the accuracy of the WTP estimates and product configuration recommendations.

References

- Adomavicius, G., and Tuzhilin, A. (2005). Towards the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extensions, *IEEE Transactions on Knowledge & Data Engineering*, 17(6), 734-749.
- Aksoy, L., Bloom, P.N., Lurie, N.H., and Cooil, B. (2006). Should Recommendation Agents Think Like People?, *Journal of Service Research*, 8(4), 297-315.
- Andrews, R. L., and Srinivasan, T.C. (1995). Studying Consideration Effects in Empirical Choice Models Using Scanner Panel Data, *Journal of Marketing Research*, 32(1), 30-41.
- Ansari, A., Essegai, S., and Kohli, R. (2000). Internet Recommendation Systems, *Journal of Marketing Research*, 37(3), 363-375.
- Bettman, J.R., Luce, M.F., and Payne, J.W. (1998). Constructive Consumer Choice Processes, *Journal of Consumer Research*, 25(3), 187-217.
- Burke, R. (2002). Hybrid Recommender Systems: Survey and Experiments, *User Modeling and User-Adapted Interaction*, 12(4), 331-370.
- Butler, J., Morrice, D.J., and Mullarkey, P. (2001). A Multiple Attribute Utility Theory Approach to Ranking and Selection, *Management Science*, 47(6), 800-816.
- Chernev, A. (2003). When More is Less and Less is More: The Role of Ideal Point Availability and Assortment in Consumer Choice, *Journal of Consumer Research*, 30(2), 170-183.
- Darley, W.K., and Smith, R.E. (1995). Gender Differences in Information Processing Strategies: An Empirical Test of the Selectivity Model in Advertising Response, *Journal of Advertising*, 24(1), 41-56.
- De Bruyn, A., Liechty, J.C., Huizingh, E.K.R.E., and Lilien, G.L. (2008). Offering Online Recommendations with Minimum Customer Input Through Conjoint-Based Decision Aids, *Marketing Science*, 27(3), 443-460.
- Diehl, K., Kornish, L.J., and Lynch, J.G. (2003). Smart Agents: When Lower Search Costs for Quality Information Increase Price Sensitivity, *Journal of Consumer Research*, 30(1), 56-71.
- Farag, N., Smith, M., and Krishnan, M. (2003). The Consumer Online Purchase Decision: A Model of Consideration Set Formation and Buyer Conversion Rate Across Market Leaders and Market Followers, *Proceedings of the ICIS 2003*, Seattle, USA.
- Gensler, S., Hinz, O., Skiera, B., and Theysohn, S. (2012). Willingness-to-pay Estimation with Choice-based Conjoint Analysis: Addressing Extreme Response Behavior with Individually Adapted Designs, *European Journal of Operational Research*, 219(2), 368-378.
- Gilbride, T., and Allenby, G. (2004). A Choice Model with Conjunctive, Disjunctive, and Compensatory Screening Rules, *Marketing Sciences*, 23(3), 391-406.
- Green, P.E., Krieger, A.M., and Agarwal, M.K. (1993). A Cross Validation Test of Four Models for Quantifying Multiattribute Preferences, *Marketing Letters*, 4(4), 369-380.
- Gu, B., Konana, P., and Chen, H. (2011). Consumer Product Consideration and Choice at Purchase Time at Online Retailers, *Proceedings of the ICIS 2011*, Shanghai, China.
- Guest, L. (1964). Brand Loyalty Revisited: A Twenty Year Report, *Journal of Applied Psychology*, 48, 93-99.
- Häubl, G., and Trifts, V. (2000). Consumer Decision Making in Online Shopping Environments: The Effects of Interactive Decision Aids, *Marketing Science*, 19(1), 4-21.
- Hauser, J.R., and Wernerfelt, B. (1990). An Evaluation Cost Model of Consideration Sets, *Journal of Consumer Research*, 16(4), 393-408.
- Helmig, B. (1997). *Variety-seeking behaviour im Konsumgüterbereich*, Wiesbaden, Gabler.
- Herlocker, J.L., Konstan, J.A., Terveen, L.G., and Riedl, J.T. (2004). Evaluating Collaborative Filtering Recommender Systems, *ACM Transactions on Information Systems*, 22(1), 5-53.
- Huang, S. (2011). Designing Utility-based Recommender Systems for E-Commerce: Evaluation of Preference-elicitation Methods, *Electronic Commerce Research and Applications*, 10(4), 398-407.
- Jedidi, K., Kohli, R., and DeSarbo, W. (1996). Consideration Sets in Conjoint Analysis, *Journal of Marketing Research*, 33(3), 364-372.

- Jedidi, K., and Zhang, Z.J. (2002). Augmenting Conjoint Analysis to Estimate Consumer Reservation Price, *Management Science*, 48(10), 1350-1368.
- Johnson, E.J., and Payne, J.W. (1985). Effort and Accuracy in Choice, *Management Science*, 31(4), 395-414.
- Karjaluoto, H., Karvonen, J., Kesti, M., Koivumäki, T., Manninen, M., Pakola, J., Ristola, A., and Salo, J. (2005). Factors Affecting Consumer Choice of Mobile Phones: Two Studies from Finland, *Journal of Euromarketing*, 14(3), 59-82.
- Lin, F. -R., Huang, S. -L., and Yang, Y. -C (2005). Using radial basis function networks to model multi-attribute utility functions, *Proceedings of the 4th Workshop on e-Business*, Las Vegas, USA.
- Miller, K.M., Hofstetter, R., Krohmer, H., and Zhang, Z.J. (2011). How Should Consumers' Willingness to Pay Be Measured? An Empirical Comparison of State-of-the-Art Approaches, *Journal of Marketing Research*, 48(1), 172-184.
- Mitchell, V.-W., and Walsh, G. (2004). Gender Differences in German Consumer Decision-making Styles, *Journal of Consumer Behaviour*, 3(4), 331-346.
- Moore, W.L. (2004). A Cross-Validity Comparison of Rating-based and Choice-based Conjoint Analysis Models, *International Journal of Research in Marketing*, 21(3), 299-312.
- Pfeiffer, J. (2012). *Interactive Decision Aids in E-Commerce*, Springer, Heidelberg.
- Resnick, P., and Varian, H.R. (1997). Recommender Systems, *Communications of the ACM*, 40(3), 56-58.
- Sambandam, R., and Lord, K.R. (1995). Switching Behavior in Automobile Markets: A Consideration-Sets Model, *Journal of the Academy of Marketing Science*, 23(1), 57-65.
- Schafer, J.B., Konstan, J.A., and Riedl, J.T. (2001). E-Commerce Recommendation Applications, *Data Mining Knowledge Discovery*, 5(1), 115-153.
- Scholz, M. and Dorner, V. (2012). Estimating Optimal Recommendation Set Sizes for Individual Consumers, *Proceedings of the ICIS 2012*, Orlando, USA.
- Tanner, W., and Swets, J. (1954). A Decision-making Theory of Visual Detection, *Psychological Review*, 61(6), pp. 401-409.
- Wallenius, J., Dyer, J.S., Fishburn, P.C., Steuer, R.E., Zionts, S. and Deb, K. (2008). Multiple Criteria Decision Making, *Multiattribute Utility Theory: Recent Accomplishments and What Lies Ahead*, *Management Science*, 54(7), 1336-1349.
- White, C.M., and Hoffrage, U. (2009). Testing the Tyranny of Too Much Choice Against the Allure of More Choice, *Psychology & Marketing*, 26(3), 280-298.
- Xiao, B., and Benbasat, I. (2007). E-Commerce Product Recommendation Agents: Use, Characteristics, and Impact, *MIS Quarterly*, 31(1), 137-209.